

An OWL Ontology for the Common Statistical Production Architecture (CSPA)

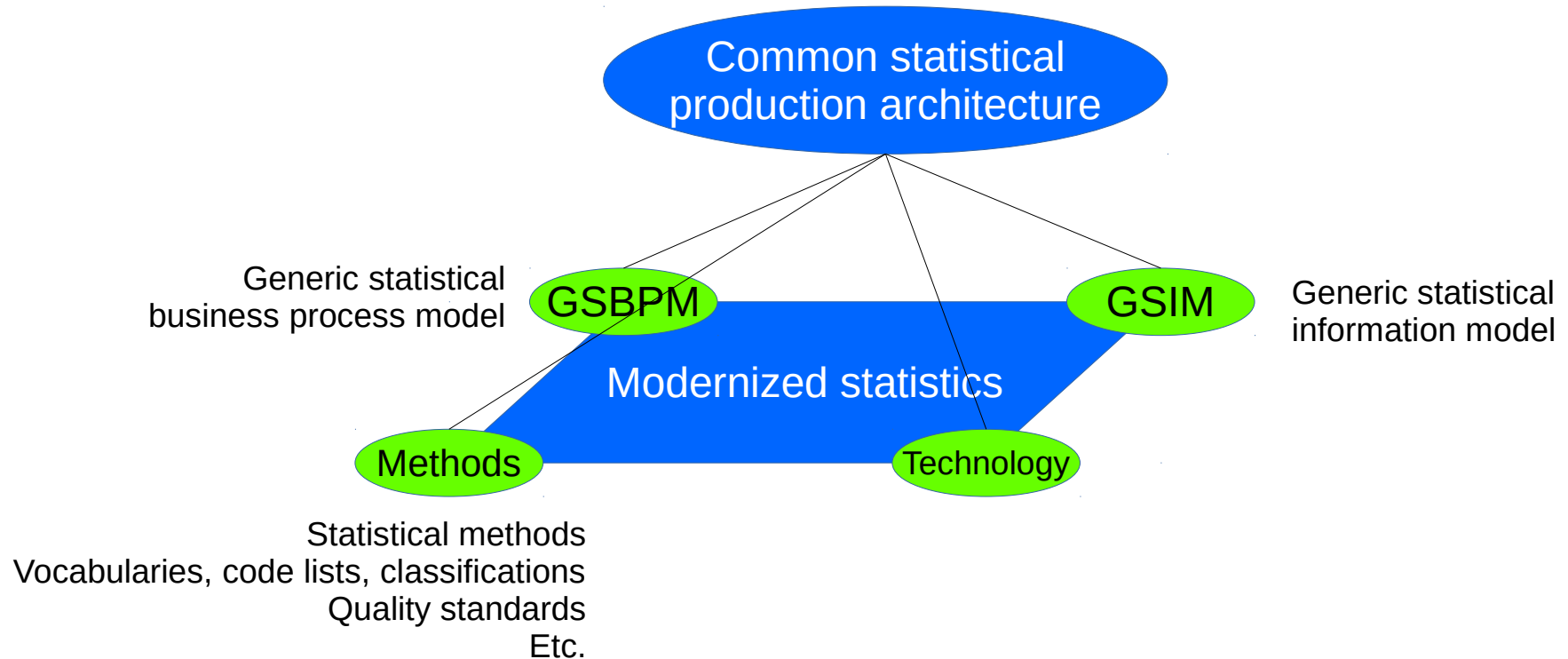
Antoine Dreyer, Franck Cotton, Guillaume Duffès
INSEE, Paris, France

The big picture

- Back in 2010, a group of NSI directors
 - Formed a high-level group (HLG-MOS) under the auspices of the UNECE (UN Economic Commission for Europe)
 - Developed a strategic paper for official statistics
 - Threats
 - internal: sclerosis on stove-pipe organizations
 - external: demand growing and diversifying, data deluge, new competitors
 - Vision
 - create conditions for innovation
 - standard-based collaboration
 - process industrialization (across domains)

The big picture

- Launched the CSPA initiative



The CSPA value proposition

- Facilitate modernization in statistical organizations
- Apply a consistent enterprise architecture approach within and across statistical organizations
- Enable international collaboration initiatives for building common solutions and services
- Foster alignment on industry standards

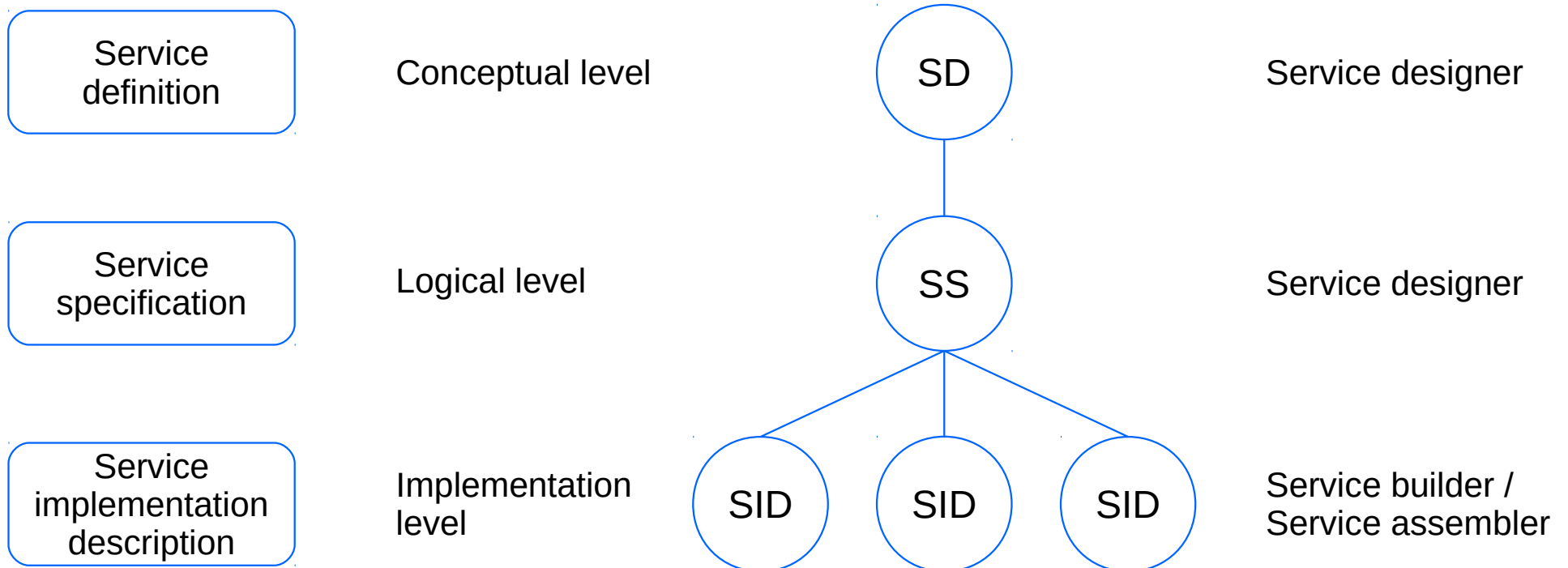
What is CSPA?

- A reference architecture for the statistical industry
- Aiming to develop **services** so they can be shared and reused
- Focused on statistical production
- Based on TOGAF

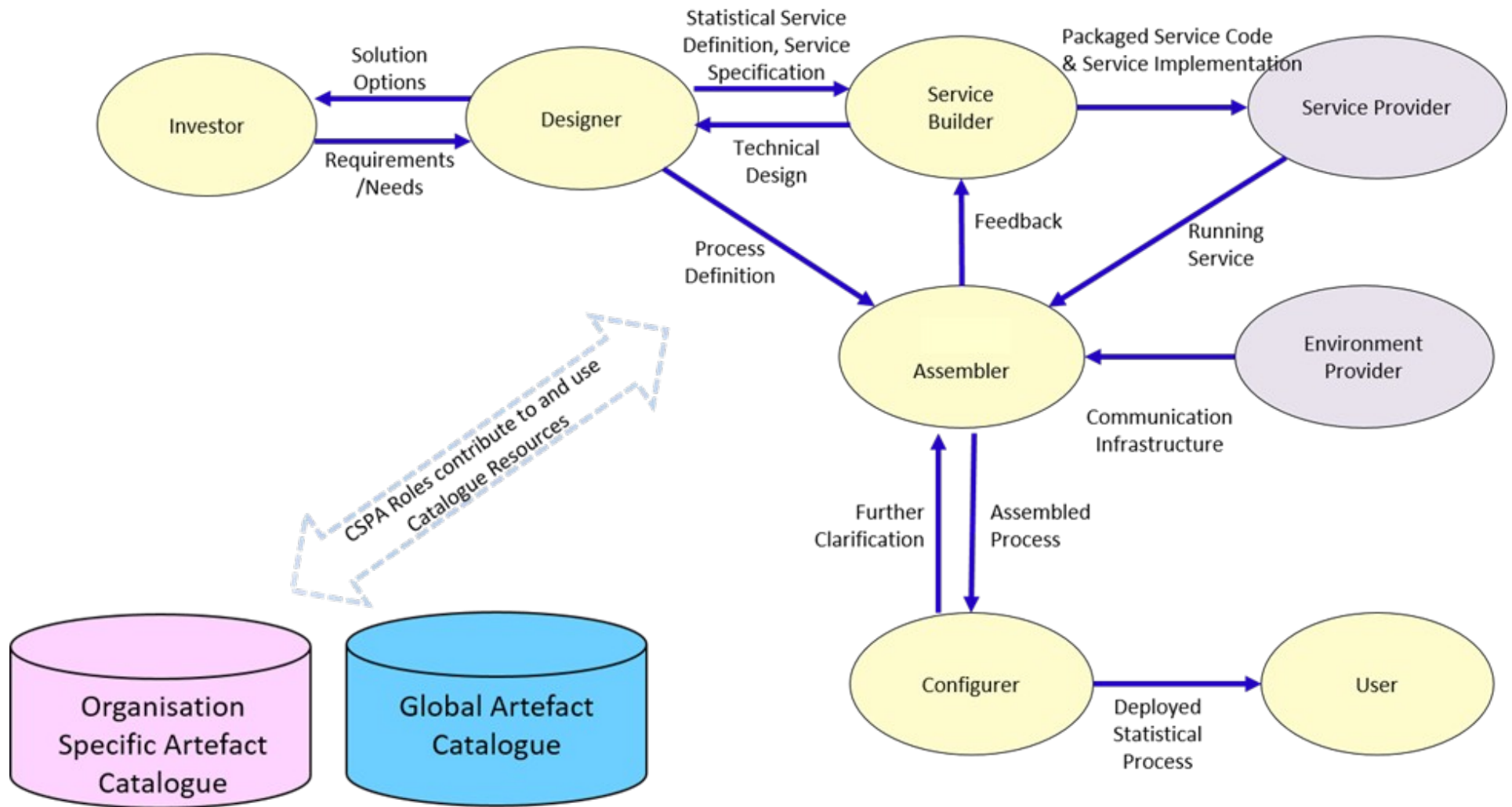
What is in CSPA?

- Business architecture principles
- Information architecture framework (introducing LIM) and principles
- Governance
- Architectural guidelines (SOA)
- **Specification of statistical services**
 - How to document a service (including templates)
 - Link to GSBPM and information models
 - Service lifecycle and associated role

CSPA services: levels of description



CSPA: roles and tools

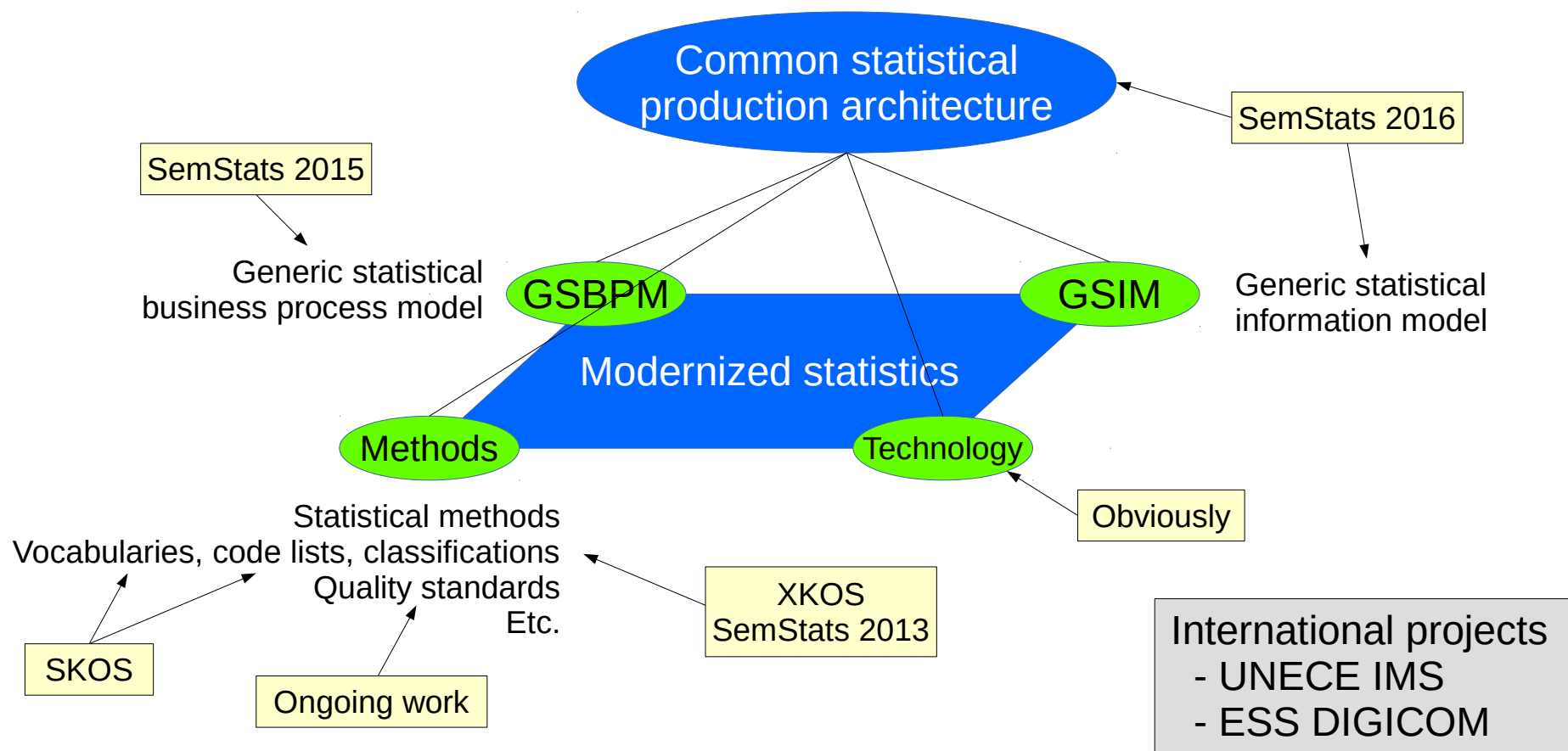


So what does it all have to do with Semstats?

- Since 2010 (and before), a group of semantic web enthusiasts works to promote SW in official statistics
 - Following the examples of DataCube, DCAT...
 - Concentrating on models, business standards, metadata
- Why?
 - Harmonize specifications formats (currently Word or PDF documents, some UML models)
 - Improve formalization and consistency within and across models
 - Publish, activate and link models and metadata

The big picture (revamped)

Semantizing CSPA



Building the CSPA ontology

- Approach followed
 - Read the specification!
 - Studied existing examples
- Identified three main semantic axes
 - Level of service description
 - Description properties (grouped by topics)
 - Added: distinction function/package (~granularity)
- Other aspect: roles in service lifecycle

Building the CSPA ontology

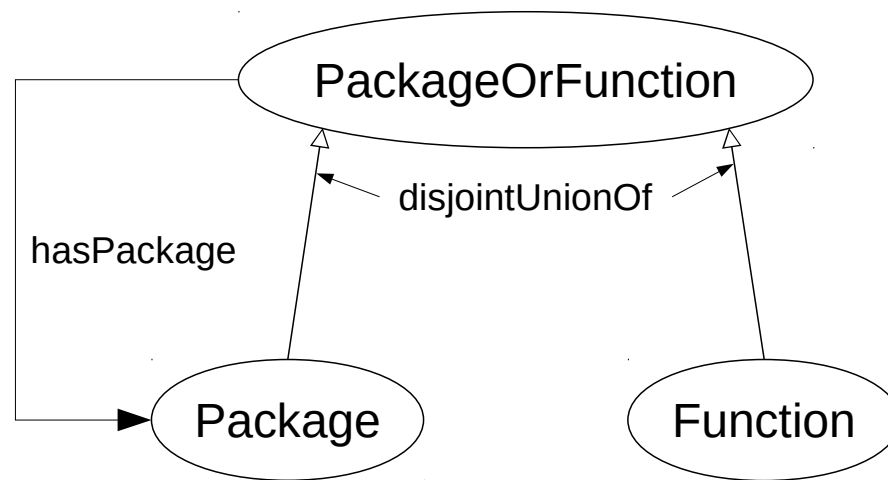
- 8 Topics
 - Identification: name, version
 - Business function: what the service is for
 - Documentation: e.g. how to use the service
 - Provenance: who designed, built, implemented the service
 - Interface: how to invoke the service
 - Dependencies: what is needed to invoke the service (OS, DBMS, statistical software...)
 - Inputs and outputs

Building the CSPA ontology

- 3 levels of documentation, 8 topics, 2 granularity categories (package/function)...
- ...but not every combination makes sense
 - Business function: conceptual level
 - Inputs and outputs only for functions, but take different forms at each level
- Some refinements
 - Dependencies can be linked to an Interface at implementation level

Ontology overview

- Granularity categories

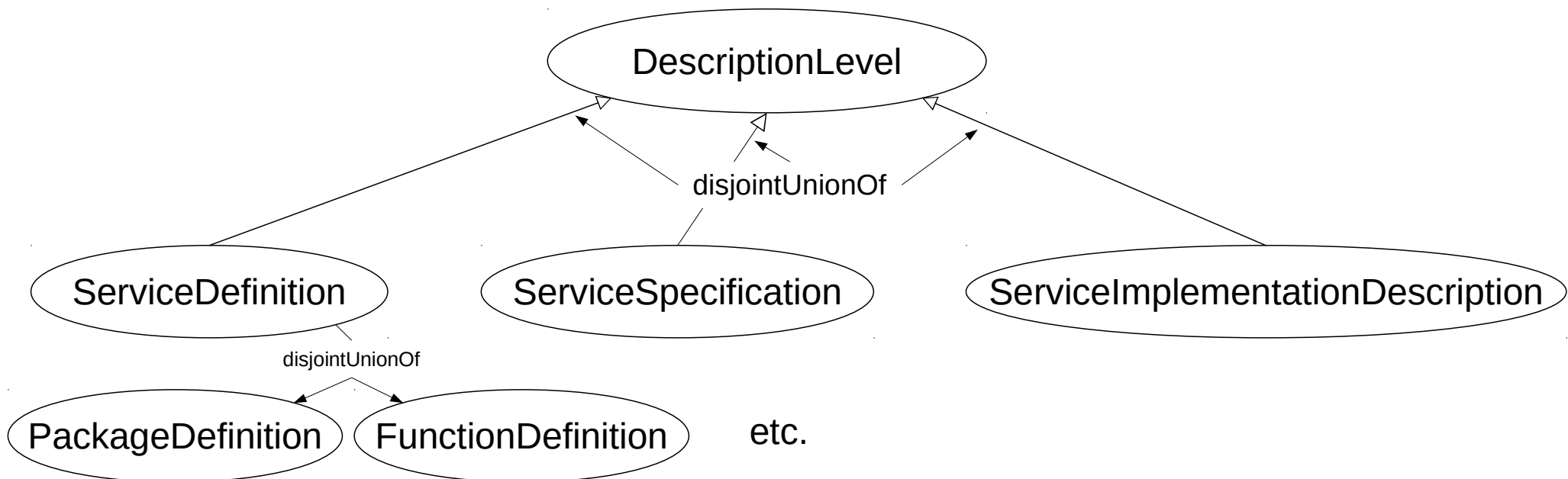


- Topics

- One class for each topic, daughter of PropertyTopic
- Then link documentation properties to topics
- Exception: identification properties linked directly

Ontology overview

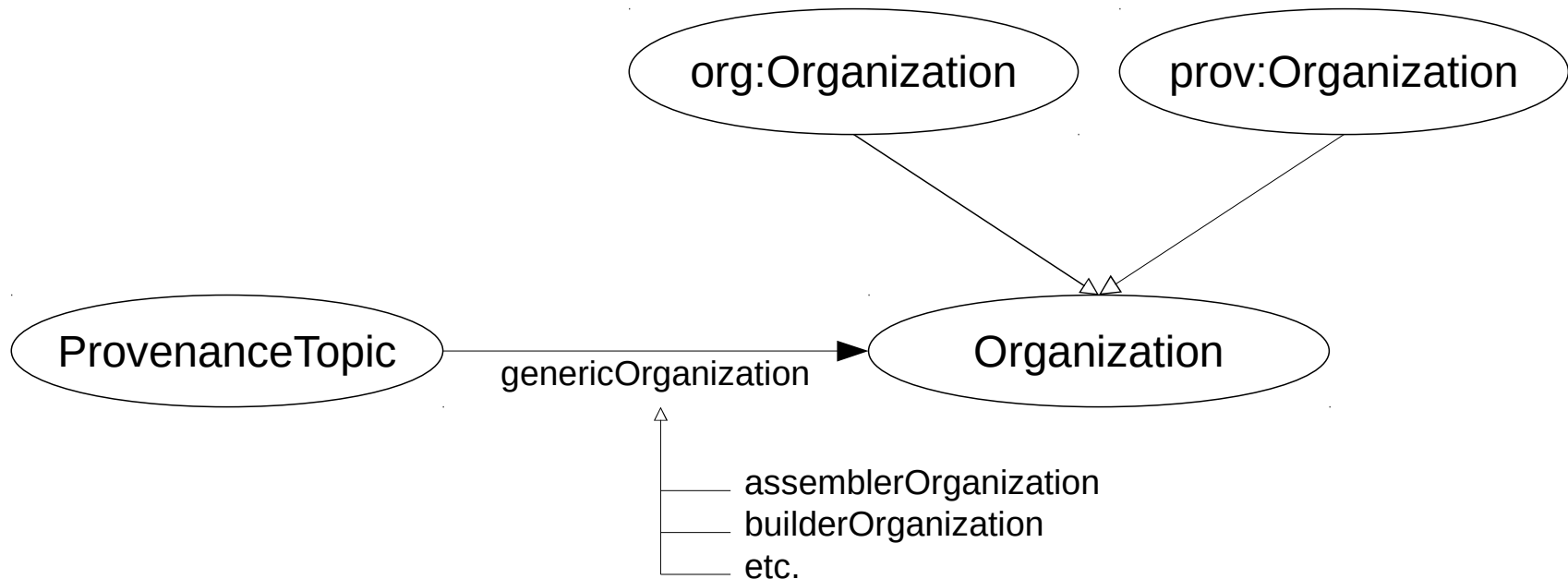
- Levels of description



- Properties to link categories to levels of description
- Properties to link topics to levels of description at the appropriate depth in the hierarchy

Ontology overview


- Roles and service lifecycle



Use of other ontologies

- The CSPA service business function topic is linked to the GSBPM SubProcess class
- The input and output classes at service definition level are linked to the GSIM object class
- Classes and properties from FOAF, PROV, DCTerms and ORG are used directly

Evaluation

-  Work in progress
- We were able to “reengineer” existing CSPA service documentation
- We will use the ontology on services currently under development
- The ontology was used to build a prototype of a CSPA service editor

Conclusion & future work

- Conclusion
 - First try at representing main parts of CSPA semantics as OWL
 - Method used is domain-specific, but can be adapted in other contexts
 - Part of a general effort for promoting semantic web approach for statistical standards
- Next steps
 - Further work on provenance aspects
 - Continue CSPA service editor development
 - Present the ontology and editor to the CSPA implementation group and the HLG

Thank you

Questions?